INF 2178 - Experimental Design For Data Science

Technical Assignment 4

By

Saad Umar

Course Instructor: Shion Guha

Date: April 05, 2024

1. Introduction

In this paper, we explore the impact of different factors on cognitive performance, specifically focussing on Estimated Total Intracranial Volume (eTIV). eTIV is a measure of the total volume inside the skull, including the brain, cerebrospinal fluid, and other structures. In Alzheimer's research, eTIV is often used to normalize brain imaging data to account for individual differences in brain size. This normalization helps researchers identify structural changes and better understand the progression of Alzheimer's disease across different individuals. The longitudinal aspect of the dataset allows for the observation of neurodegenerative progression over time, offering insights into the disease's trajectory and the effectiveness of potential treatments. Analyzing this data not only aids in discovering the complexities of Alzheimer's disease but also assists by tailoring interventions to individual aging patterns and disease stages.

1. **Research Question 1:** Does the 'Group' variable have a significant effect on eTIV across different visits?
2. **Research Question 2:** Is there a significant difference in eTIV between males and females across different visits?

By addressing these questions, we aim to contribute insights into dynamics of Alzheimers and provide a deeper understanding that can potentially inform more effective interventions.

2. Data Cleaning & Data Wrangling

The raw dataset has a total of 16 columns with 294 entities (rows). After initial review of the dataset, I was confident that not much data cleaning was necessary for the scope of my analysis. However, I removed one feature named '*Unnamed: 0'* while I kept all other features which might be necessary for future analysis, however, I did not use all of them for this specific analysis. Below I have outlined some of my observations:

a. Observations & Considerations

My analysis is quantitative and I've only used specific columns from the dataset to perform the analysis. Below I have provided a short description of each columns that I have used in my analysis so far:

- Subject ID: ID of the subject (patient)
- Group: Patient categorized as (Demented, non-demented, converted)
- Visit: Number of visits
- M/F: Gender of the patient (Male/Female)
- eTIV: Estimated Total Intracranial Volume - a metric (related to Alzheimers)

3. Exploratory Data Analysis (EDA)

I performed an extensive EDA to leverage insight that could potentially help me derive insightful research questions. I started by summarizing quantitative data and then using a bunch of bar plots and boxplots to see how different features varied and how the distributions differed across different groups.

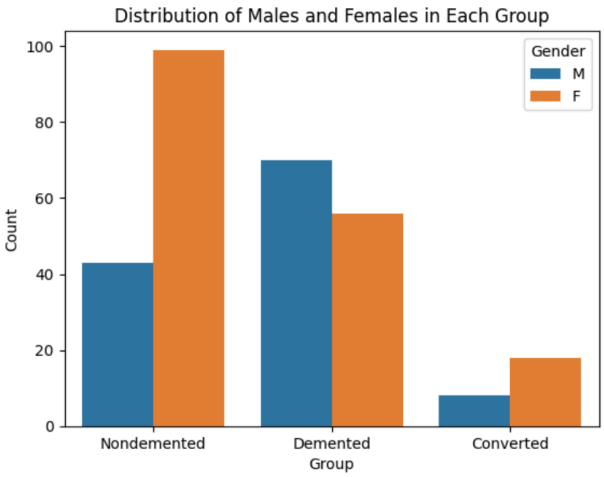

Fig 1 - Distribution by gender

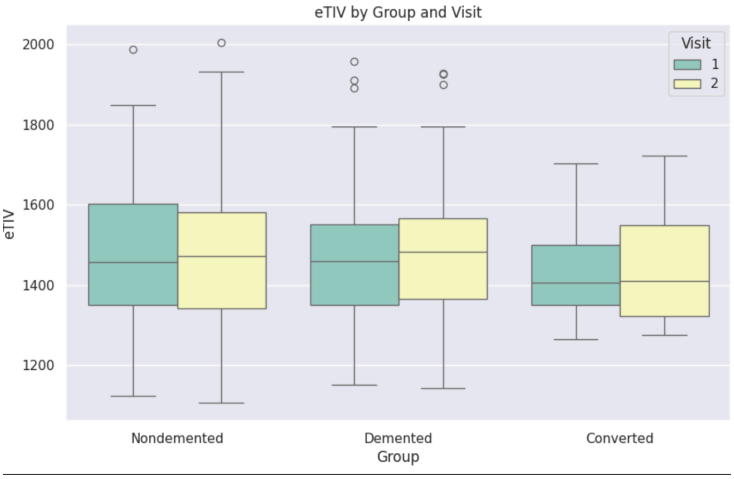

Fig 2 - Distribution of eTIV by group & Visit

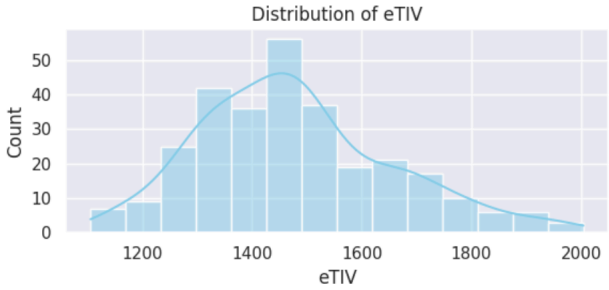

Fig 3 - Distribution of eTIV

4. Mixed Anova 1

Research Question 1: Does the 'Group' variable have a significant effect on eTIV across different visits?

I used a two-way mixed effect anova to test the hypothesis, *there is no significant difference in eTIV scores among the different groups across visits.* While my alternative hypothesis being: *there is a significant difference in eTIV score among the different groups across visits.* Table 1 presents my results. We can see that the

	Mean Sq (MS)	p-unc
Group	18712	0.743
Visit	5574	0.003
Interaction	502	0.438

Table 1 - MixedEffect Anova 1

between-subjects factor (Group) has a p-value > 0.05, thus, we fail to conclude that Group has a statistically significant effect on the eTIV score. However, *Visit* has a p-value of 0.003 which is < 0.05, this has a significant effect on eTIV. Following from theory, since one of the factors is not significant, thus the interaction term is not significant either (in addition to its p-value being > 0.05).

Group	W	p-val
Non demented	0.973	0.006
demented	0.968	0.005
Converted	0.918	0.041

Table 2 - Shapiro-Wilk Test

Visit	W	p-val
1	0.94	0.39
2	0.41	0.67

Table 3 - Levene's Test

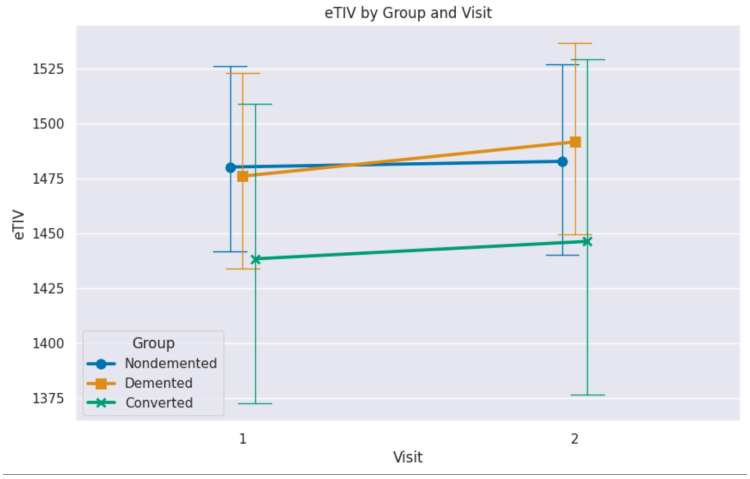

Fig 4 - Point Plot for RQ1 by Group & Visit

The results in table 4 (from post-hoc test) suggest significant differences in some comparisons (especially in visit over time for the paired samples) but not in others, particularly when comparing groups or the interaction between visit and group. (Note, not all results are shown because of space constraints). Fig 4 shows: the demented and converted group shows a slight increase in eTIV on the second visit, whereas the NonDemented group almost remains the same.

Contrast	Visit	(A,B)	p-unc
Visit*Group	1	(Converted,Demented)	0.40
Visit*Group	1	(Converted, NonDemented)	0.41
Visit*Group	2	(Converted,Demented)	0.367

*Table 4 - Post-hoc tests1 - (I have only shown limited values) because of **space constraint***

5. Mixed Anova 2

Research Question 2: Is there a significant difference in eTIV between males and females across different visits?

	Mean Sq (MS)	p-unc
M/F	3023632	<0.001
Visit	5574	0.003
Interaction	238	0.531

Table 5 - MixedEffect Anova 2

I used a two-way mixed effect anova to test the hypothesis, *there is no significant difference in eTIV scores among the different genders across visits.* While my alternative hypothesis being: *there is a significant difference in eTIV score among the different genders across visits.* Table 5 presents my results. We can see that the between-subjects factor (Gender) has a p-value < 0.05, thus, we can say that Gender has a statistically significant effect on the eTIV score. Additionally, *Visit* has a p-value of 0.003 which is < 0.05, this also has a significant effect on eTIV. However, the interaction term is not significant, as its p-value is > 0.05.

M/F	W	p-val	normal
M	0.979	0.053	True
F	0.987	0.125	True

Table 6 - Shapiro Wilk Test For RQ 2

Visit	W	p-val
1	0.94	0.39
2	0.41	0.67

Table 7 - Levene's Test for RQ 2

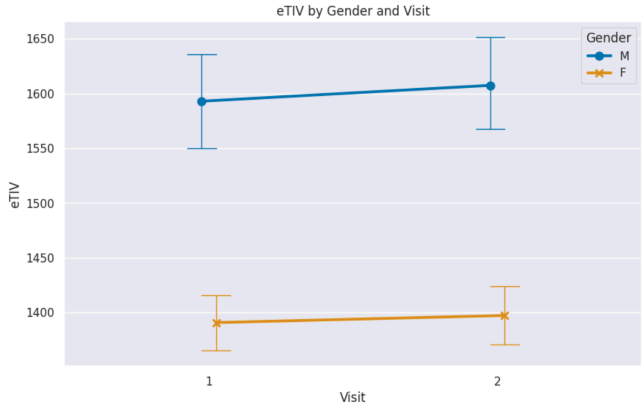

Fig 5 - Point Plot for RQ2 by Gender & Visit

The results in table 8 (from post-hoc test) suggest significant differences in most comparisons (Note, not all results are shown because of space constraints). Fig 5 shows that both male and females have slight increases (males have greater increase) in eTIV on their second visits. Generally, males have a higher eTIV than females.

Contrast	Visit	(A,B)	p-unc
Visit*M/F	1	(F,M)	<0.001
Visit*M/F	2	(F,M)	<0.001

*Table 8 - Post-hoc tests2 - (I have only shown limited values) because of **space constraint***

6. Power Analysis

I conducted a power analysis to determine the appropriate sample size in order to achieve a power of 0.91, given an effect size of 0.7 and alpha of 0.05. The **computer sample size was 45.** The power curve is presented in figure 6.

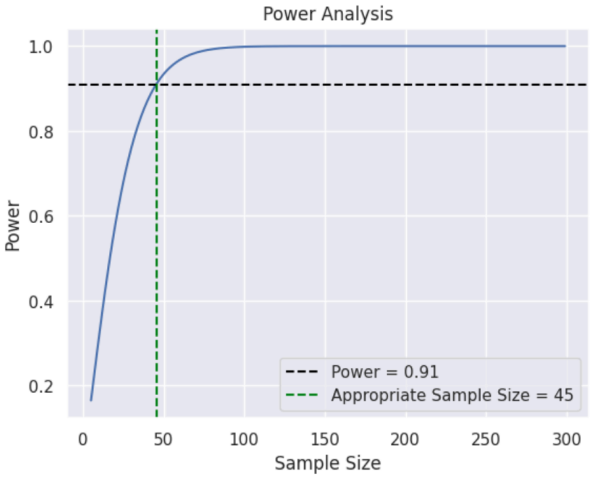

Fig 6- Power Analysis Curve

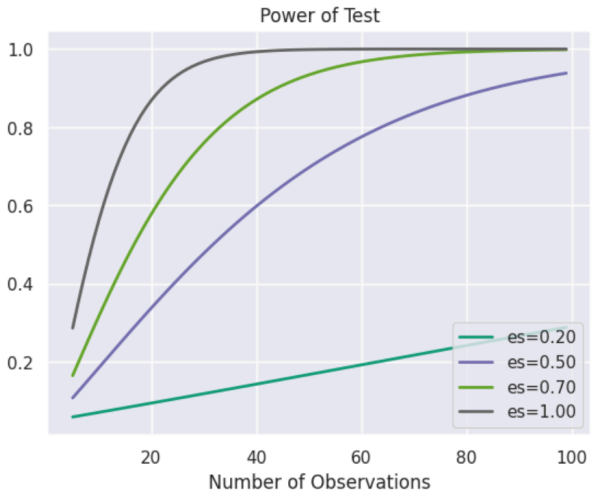

Fig 7 - Power test with different Effect Sizes

7. Conclusion

In this study, we sought to explore the impact of different factors on cognitive performance, specifically focusing on Estimated Total Intracranial Volume (eTIV), within the context of Alzheimer's research. Our analysis revealed that while the 'Group' variable did not exhibit a statistically significant effect on eTIV across different visits, the 'Visit' factor itself and 'Gender' did. Specifically, we found that both the number of visits and gender significantly influence eTIV measurements, with males showing a higher increase in eTIV on subsequent visits compared to females. These findings underscore the importance of considering both temporal and gender-specific differences when analyzing brain volume in Alzheimer's disease research. Through a combination of exploratory data analysis, mixed-effect ANOVA, and post-hoc testing, this study contributes valuable insights into the dynamics of brain volume changes in Alzheimer's patients. It highlights the need for tailored interventions that account for these variations, ultimately paving the way for more personalized and effective treatments in combating this neurodegenerative disease.